Results

The smoking rates across Pennsylvania counties show a moderate range from 0.1820 to 0.2790, with most counties having rates between 0.2260 and 0.2565. The mean smoking rate is 0.2376, slightly higher than the median (0.2300), suggesting a slight right skew. This indicates that while most counties have relatively low smoking rates, a few counties with higher rates pull the overall average up.

This trend is further highlighted in Figure 1, where the histogram displays the positively skewed distribution of county-specific smoking rates in Pennsylvania, with most counties showing lower smoking rates. Figure 2, presenting the histogram of lagged smoking rates, shifts slightly to the left compared to the observed distribution, indicating that lagged smoking rates are generally a bit lower. This suggests that, on average, counties in Pennsylvania have slightly higher smoking rates than their neighboring counties. The lower mode of the lagged distribution, relative to observed smoking rates, reinforces this pattern. Similarly, Figure 4, which maps the lagged smoking rates, aligns with the histogram findings; it is visibly less intense than Figure 3, which maps the observed smoking rates. The reduced brightness in Figure 4 suggests that lagged values are lower than observed smoking rates, further supporting the idea that, overall, counties tend to have somewhat higher smoking rates than their neighbors.

Figure1:County-specific smoking rates in Pennsylvania

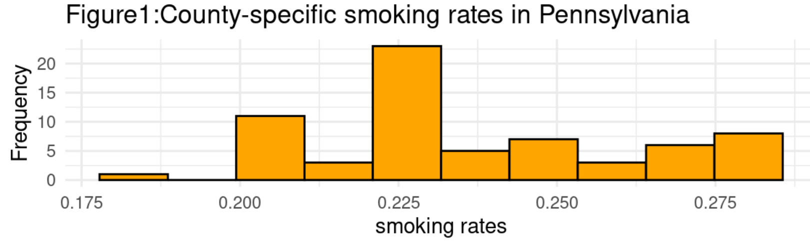


Figure2:Distribution of lagged county-specific smoking rates in Pen

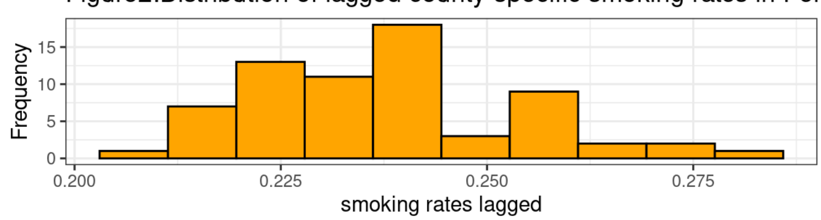


Figure3:County-Specific Smoking Rates

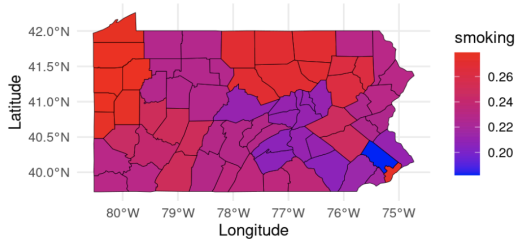
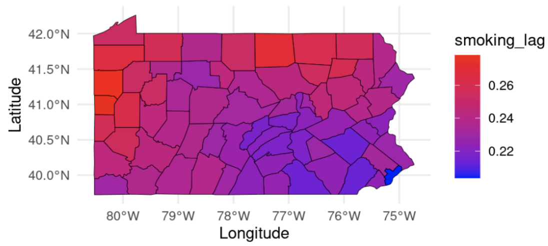


Figure4:County-Specific Smoking Rates Lagged



To determine whether smoking rates across Pennsylvania counties exhibit significant spatial clustering, we applied the Global Moran's I test. The null hypothesis (H_0) posits no spatial autocorrelation, meaning any observed clustering would be due to random variation, while the alternative hypothesis (H_a) suggests positive spatial autocorrelation, indicating that counties with similar smoking rates tend to cluster or dispersion. Using a significance level of 0.05, we calculated a Global Moran's I statistic of 0.4044. With a p-value of 3.022e-08, which is far below our threshold, we rejected the null hypothesis in favor of the alternative. These findings strongly suggest positive spatial autocorrelation among smoking rates across Pennsylvania counties, meaning that counties with similar smoking rates are geographically closer to each other than would be expected by chance.

Moving to the analysis of spatial autocorrelation, Figure 5 which is based on the Global Moran's I, offers insights into the clusters, with numerous points in the High-High and Low-Low quadrants, which support the presence of positive spatial autocorrelation. A few points in this figure also show negative autocorrelation. However, the map in Figure 7, which is based on the simulated p-values from the Local Moran's I test, indicates that many counties do not exhibit statistically significant spatial autocorrelation. This suggests that smoking rates across most counties are not clustered or dispersed significantly.

Figure5:Spatially Lagged Crime vs Crime (Scaled)

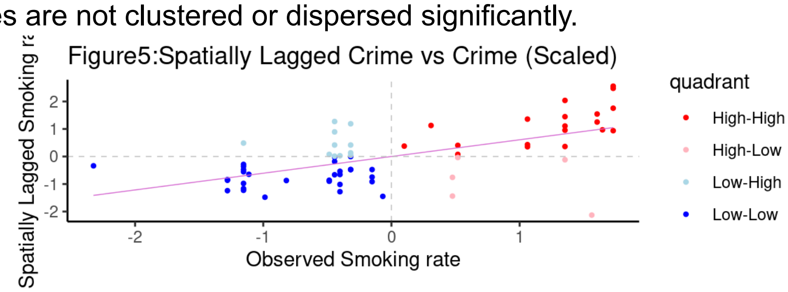
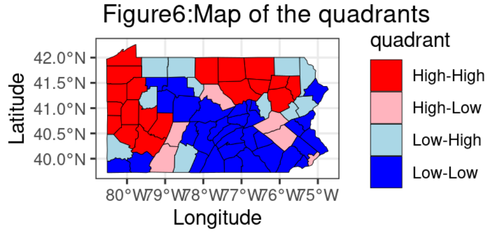


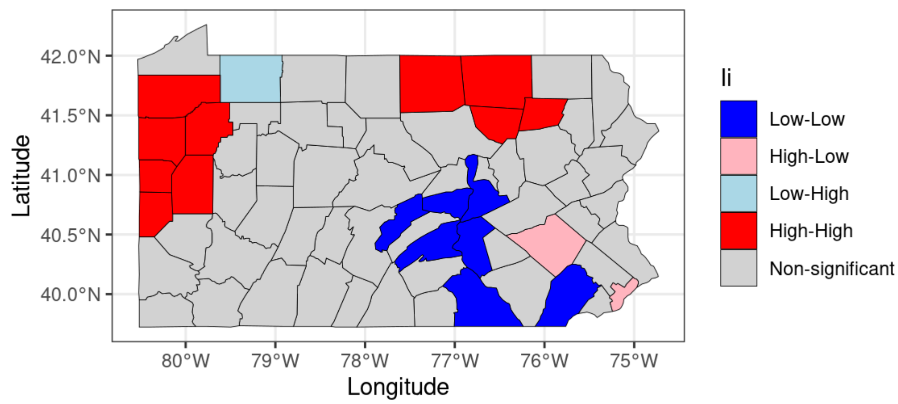
Figure6:Map of the quadrants



The clustering patterns in Figure 7 also reveal distinct high and low smoking rate clusters. Specifically, low smoking rate clusters appear in central and southeastern regions, while high-rate clusters are prominent in the northwestern and northeastern parts of Pennsylvania. These clusters reinforce the evidence of positive spatial autocorrelation. A few counties, however, show significant negative autocorrelation. For instance, one county in the northwest has a low smoking rate surrounded by high-rate counties, and two southeastern counties with high smoking rates are surrounded by low-rate counties.

The clusters observed in Figure 7 align with the patterns in Figure 6, where high-rate counties cluster with other high-rate counties and low-rate counties cluster together. Nonetheless, many low-rate counties in central Pennsylvania are non-significant in Figure 7, despite clustering in Figure 6.

Figure7:Simulated P-value(alpha=0.05)



To illustrate specific cases, Philadelphia and Cameron counties(the most and least populous counties) and their neighboring counties are plotted in Figure 8.

Figure8:Philadelphia and Cameron Counties and Neighbors

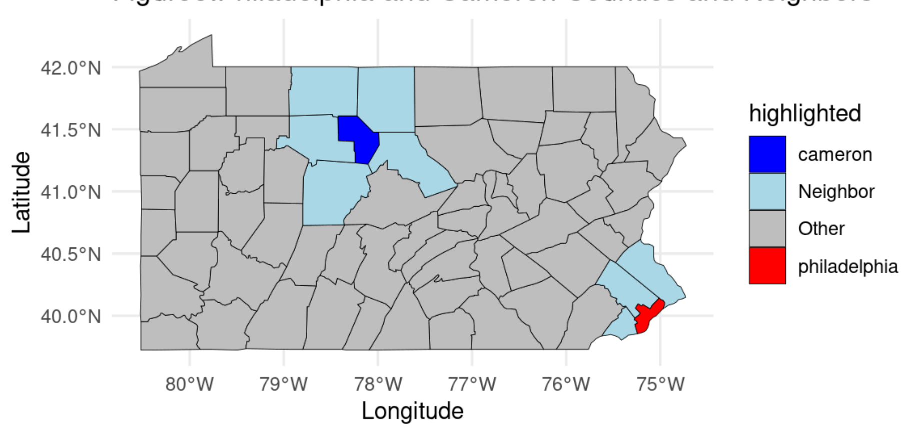


Figure3 shows Cameron county, with a smoking rate between 0.22 and 0.24, is surrounded by counties with higher smoking rates (above 0.26) or similar rates which is also reflected in Figure 6 as it places Cameron on the Low-High quadrant, indicating negative spatial autocorrelation. Yet, Figure 7 shows no significant clustering for Cameron, suggesting that its smoking rate does not create statistically significant clusters with nearby areas. In contrast, Philadelphia, with a high smoking rate (above 0.26), is surrounded by counties with lower rates, situating it in the Low-High quadrant in Figure 6 and showing significant negative spatial autocorrelation in Figure 7, which underscores its distinct smoking rate from neighboring counties.

A sensitivity analysis (Figures 9 and 10) examines two additional significance levels (0.1 and 0.01), confirming strong evidence of spatial autocorrelation across Pennsylvania. Figure 9 which shows the simulated p-value of the local Moran's I, based on an alpha level of 0.1, resembles Figure 7 but highlights slightly more areas with significant positive and negative clustering. On the other hand, Figure 10, which shows the simulated p-value of the local Moran's I, based on an alpha level of 0.01, shows mostly non-significant clustering or dispersion across counties, with the exception of a high-rate cluster in the northwest and a central-northern county displaying positive autocorrelation. Negative spatial autocorrelation appears in two southern counties with higher rates than their neighbors.

Figure9:Simulated P-value(alpha=0.1)

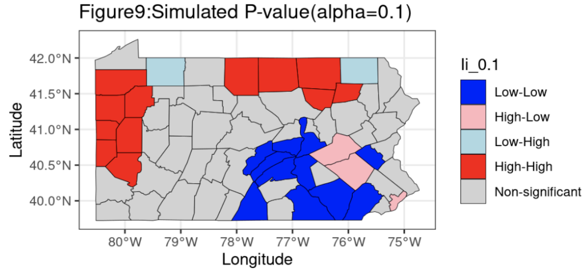
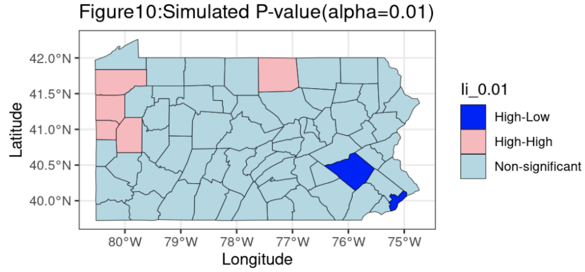


Figure10:Simulated P-value(alpha=0.01)



Finally, to adjust for multiple comparisons, a false discovery rate (FDR) correction was applied to the Local Moran's I, with Figures 11 through 13 displaying the maps of the simulated p-values. At alpha levels of 0.01 and 0.05, no significant spatial autocorrelation appears across the state (Figures 11 and 13). However, at an alpha of 0.1 (Figure 12), some areas still exhibit significant spatial autocorrelation, making it a balanced choice for identifying meaningful clusters.

Figure11:Simulated P-value(0.1), fdr adjusted

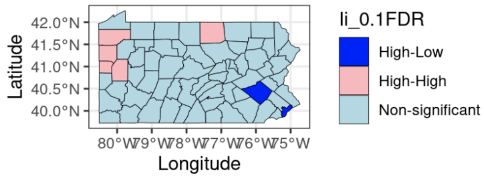


Figure12:Simulated P-value(alpha=0.01), fdr Adjusted

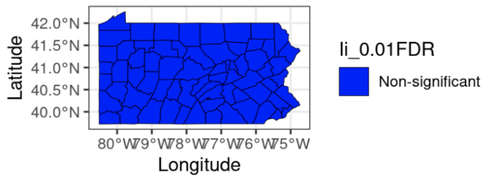
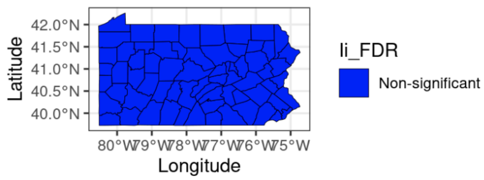


Figure13:Simulated P-value(alpha=0.05), fdr Adjusted



Discussion

The findings reveal significant spatial autocorrelation in smoking rates across Pennsylvania counties, suggesting that smoking behavior is not randomly distributed but tends to cluster geographically. High-high clusters (counties with high smoking rates near other high-rate counties) are prominent in the northwestern and northeastern regions, while low-low clusters (low-rate counties near other low-rate counties) appear in central and southeastern Pennsylvania.

It is important to note that across these sensitivity levels, Philadelphia consistently shows significant negative spatial autocorrelation, while Cameron County has no significant spatial autocorrelation with its neighbors. This discrepancy could be influenced by differences in population sizes.

This clustering could also be due to regional cultural or socioeconomic factors that influence smoking behavior, such as public health policies, or community norms. Importantly, Philadelphia shows significant negative spatial autocorrelation, indicating higher smoking rates compared to neighboring counties with lower rates. The consistent pattern seen in Philadelphia may reflect its high population density, which could amplify social and environmental factors impacting smoking prevalence.

One limitation is that the local Moran's I results show non-significance for many counties, indicating that spatial clustering is only partially present across the state. This could be attributed to the use of county-level data, which may overlook finer, neighborhood-level patterns of smoking behavior. Additionally, the sensitivity of the findings to different significance thresholds (0.01, 0.05, 0.1) highlights the challenge of choosing an appropriate level of significance for spatial analyses. Adjusting for false discoveries (FDR) further reduces the number of significant clusters, which might limit the Department of Health's ability to target specific high-risk areas.

To address these limitations, a multiscale approach could be adopted. By incorporating finer-grained geographic data, such as census tracts or neighborhoods, we could identify smaller hot-spots within counties and gain a clearer understanding of localized smoking patterns.

Given the identified clusters of high smoking rates, particularly in the northwestern, northeastern and Philadelphia counties, the Pennsylvania Health Department should prioritize these areas for smoking reduction campaigns. Initiatives could include targeted outreach and education, increased access to cessation resources, and community-based interventions. Furthermore, given the advantage of using a 0.1 significance level, it would be practical for the department to use this threshold for targeting interventions, as it balances between identifying significant areas and controlling for false positives.